**“The impacts of gender disparity, marital status and educational qualification on the housing expenditure and financial status of a family.”**

Final research project

Habiba Aziz

DATA-710 SAS

**I. Introduction.**

Gender inequality is one of the most arguable topics in almost every society around the globe. According to the 2016 report from Inequality.org, among full-time workers, women earned less than 81 cents for every dollar a man earned across United States. There are numerous concepts presented by social scientists on disparities between genders. To study the gender gaps among householders themselves is a challenge because there are many factors contributing such as Education level, employment, family/household income, marital status, type of housing and so on. For my project I am using AHS (American Housing Survey) data from 2017 to check if there any disparity in a householder’s family income based on Gender?

Income, marital status, and education can be strong indicators of housing expenses and this will be the basis of my other research question i.e., do education, income, and marital status have any effect on housing and expenditure?

**II. Descriptive Analysis.**

II.1. Selection of 10 variables from the American Housing Survey data

The variables chosen for this project from AHS 2017 data are HHSEX (Householder sex), BLD (type of housing units), FINCP (Sum of all types of income for the householder), HINCP(Family income), HHMAR (Marital status of householder), HHGRAD (Educational level of householder), GASAMT (Monthly gas amount), UTILAMT (Monthly utility amount), HHAGE (Age of householder), and TENURE(Owner or renter status of unit). Moreover, missing values were recorded in SAS and erased before analysis, and sample of 2000 observations from the data were selected by random sampling method drawn from the population. (See **VI. Appendix II.1**)

II.2. Summary of the variables

Sex of householder (HHSEX) is utilized as an independent variable and was almost equally distributed i.e., Male 50.52%, and Female 49.48% of the population. Educational level (HHGRAD) which was further recoded as EDRP (educational group) is another independent variable and was divided into three groups: where 37.21% of householders had high school or less level education, 49.71% householders had associates or bachelors, and 13.08% house holders had masters or higher level degree (see **Figure 1. In V**.). HHMAR (Marital status) of the householder is also one of the independent variables, and it has further six categories: married and spouse present consisted of 45.70%, while married and spouse absent was 2.12%, widowed was 9.84% , divorced was 16.55%, separated was 3.26%, and never married was 22.52% of the population distribution. The last independent variable is age of householder (HHAGE) and its mean was 51.53 and standard deviation was 16.65.

Family income (FINCP), monthly total utility amount (UTILAMT), monthly gas amount (GASAMT) are the dependent variables for this project. The population mean for family income was $71955.43, monthly utility cost mean was $204.77, and monthly gas amount was $37.23. Household income (HINCP) is also a dependent variable and further recoded into four income group i.e., lowest income group was 25.26%, low income group was 25.35%, high income group was 22.74%, and highest income group was 26.64% of the population distribution. Type of housing unit (BLD) distributed in 10 groups is also a dependent variable, where the code 01 was for mobile home, 02 one-family house detached, 03 one family house attached, 04 two apartments, 05 3-4 apartments, 06 5-9 apartments, 07 10-19 apartments, 08 20-49 apartments, 09 50 or more apartments, and 10 boat, RV, or Van. The last dependent variable for this project is owner or renter status (TENURE) which was grouped into three categories; owned or being bought had 58.39%, renters had 40.05%, and occupied without payment of rent had 1.56% of the distribution.

**III. Inferential Statistics.**

III.1. Confidence intervals

Confidence interval of interval/ratio variable was monthly utility cost (UTILAMT) and for proportion it was householder sex (HHSEX). Using 95% confidence interval and alpha level 0.05 on utility cost, it was determined that there was 95% of confidence interval that population mean (204.77) fell within 200.42 and 210.66 confidence level (CL) for the mean. It can be concluded that population mean for UTILAMT fell within confidence interval. (See Appendix III.1). For the proportion variable HHSEX used 0.05 alpha level, with 95% of confidence interval the proportion of Male population (50.52%) and Female population (49.48%) which fell within 47.80% < Male/Female < 52.19% CL for both the genders. Consequently, it was further concluded that the sample drawn for this project was from same population. (See **VI. Appendix III.1**)

III.2. One- and Two-sample t-tests

A one-sample t-test was applied on the age of householder (HHAGE) variable within the sample of population. Initially all the test requirements were met, counting the interval-ratio analysis, random sample of 2000 observations were drawn from the population based on EPSEM, and the sampling distribution was considered to be in normal shape. The mean age of householder in the population was 50.53. Using an alpha level of 0.05, the obtained t value (-2.06) fell in the critical region (± 1.96), letting us to reject the null hypothesis (H0: U1=U2). With 95% of confidence interval there was a statistical significance difference between population mean and sample mean. (See **VI. Appendix III.2**)

A two-test sampling with family income (FINCP) and genders (HHSEX) as variables was observed. The test provisions were met, 2000 samples were randomly selected from AHS data based on EPSEM, the sampling was considered to be in normal shape, and if the variances of two groups were equal the pooled t-test was needed, else in case of non-equality Satterthwaite t-test was the second option. Null hypotheses were stated (H0: U1=U2) mean family income of Male householder (encoded as 1 in AHS data) was equal to Mean Family income of Female householder ( encodes as 2 encoded in AHS data) i.e., the two-population means were equal and independent (H1: U1≠ U2). The population means were not equal and there was a significant difference. Variances of both genders of family income were not equal, and thus a Satterthwaite estimate of the common variance was used. Using a 95% confidence interval and alpha level (0.05) greater than P value (.0001), the obtained t value (4.36) fell within critical region (± 1.96)**.** Therefore, null hypotheses were rejected i.e., mean family income of male respondents was statistically significantly higher than female respondents. (see Fig. 2 **In V**.)

III.3. Chi-square test

Chi-square tests between gender (HHSEX) and its relationship with owner status (TENURE) was studied. TENURE was recoded in three groups based on owned, rented and occupied without payment response codes. For the first step of hypotheses, testing all the requirements were met, including random sampling, two nominal (HHSEX and TENURE) level measurements, and sampling distribution in normal shape. The null hypotheses were stated, H0: f0=fe, i.e., the variables were independent, and H1: f0≠ fe, i.e., the variables were not independent.

The obtained chi-square value for TENURE was 11.50 which was greater than the critical value (± 5.991) and alpha value 0.05 was greater than p < 0.0032, so the null hypotheses were rejected. Based on hypotheses test of chi square, it can be said that Gender and Owner status were dependent on each other. Therefore, there was a statistically significant relationship between Gender and Owner status in the population from which the sample was drawn. (See **VI. Appendix III.3**)

III.4. Crosstabulation analysis

For this part I chose two variables HHGRAD i.e., ordinal, which was further recoded into 3 groups (EDGRP), and HINCP which was recoded into 4 groups (INGRP). (see **Table1.** and **Table2.** **In V**.)

Here, EDGRP (Education group) was an independent variable and INGRP (Income group) was a dependent variable. For Education group if the response code for years in education was than 39 then it fell under group 1 i.e., high-school or less education, if the response code for education level years were from 39 to 45 it fell under group 2, i.e., bachelors or less education , and if the response code for education level was 45 years or more it fell under group 3, i.e., for Masters, PHD etc. For Income groups, if the household income was less than or equal to 22800 then it fell under group 1, i.e., lowest household income, if the household income ranged between 22800 to 50000 then it fell under group 2 , if it ranged between 50000 to 90000, it fell under group 3, and finally, if the income was greater than 90000, it fell under group 4. The bi-variate table in the output file portrays the distribution of income group (dependent variable) and education group (independent variable). There was a slight association between the variables i.e., it does not follow the widely accepted social belief that the better the education level, the better the income level. This sample showed a trend where the education level of householders from group 2 i.e., bachelors or less had the highest share of income group, which shows that not many people consider higher education. (See **VI. Appendix III.4**)

III.5. One-way Analysis of Variance (ANOVA)

A one-way analysis of variance was used to discover the association between marital status (HHMAR) and monthly utility cost (UTILAMT). All test assumptions and requirements were met, comprising the use of independent random samples based on EPSEM, an interval-ratio level of measurement, normally distributed populations, and equal population variances. The ANOVA test delivered an F-value of 190.71, which fell within the critical region using an alpha level of 0.05 (dfb=2, dfw=1997, critical value= +- 3.31). Here as the test statistics fell in the critical region, the null hypothesis was rejected. There was a 95% statistical significance between different marital groups and their utility amount expenditure i.e., the monthly utility amount expenditure for all the 3 marital groups was significantly different from one another. Married people spent more towards their monthly utility than that of divorced, widowed, separated, and never married groups. (See **VI. Appendix III.5**)

**IV. Conclusion.**

There were couple of interesting findings in this project and statistical analysis played a very vital role in terms of the results. The gender disparity in family income was proved and statistically it was seen that there was a significant difference between male and female family incomes. Furthermore, it was observed that the ownership of males for owned houses was more than that of females, and it was observed that more female householders resided in rented spaces than that of male householders. It was also found that married people were likely to spend more towards their monthly expenditure than that of unmarried and separated/divorced/widowed groups. The relationship between educational level and income group provided the details that most of the people do not pursue higher level of education.

**V. Tables and Figures.**

***Figure 1:*** Percentage of Educational group among householders.

***Figure 2.*** Sample Mean Family and Household Income by Sex

|  |  |  |
| --- | --- | --- |
| **Table 1. Distribution of Income group in the sample** | |  |
| Income Group | Percent % | Cumulative Percent% |
| Lowest-1 | 23.45 | 23.45 |
| Low-2 | 25.00 | 48.45 |
| High-3 | 24.55 | 73.00 |
| Highest-4 | 27.00 | 100.00 |
| Total | 100.00% |  |

|  |  |  |
| --- | --- | --- |
| **Table 2. Distribution of education group in the sample** | |  |
| Education Group | Percent % | Cumulative Percent% |
| Group1 (High school or less) | 37.00 | 37.00 |
| Group2 (Bachelors or less) | 49.25 | 86.27 |
| Group3 (Maters, PHD or more) | 13.75 | 100.00 |
| Total | 100.00% |  |

**VI. Appendix: SAS-related documents.**

II.1. Selection of 10 variables from the American Housing Survey data.

**SAS PROGRAM FILE:**

libname project "\\tsclient\Habiba Aziz\Desktop\SASFINAL";

**data** one; set project.housenew;

ods rtf;

**proc** **contents**; **run**;

ods rtf close;

**SAS LOG FILE:**

78 libname project "\\tsclient\Habiba Aziz\Desktop\SASFINAL";

NOTE: Libref PROJECT was successfully assigned as follows:

Engine: V9

Physical Name: \\tsclient\Habiba Aziz\Desktop\SASFINAL

79 data one; set project.housenew;

80 ods rtf;

NOTE: Writing RTF Body file: sasrtf.rtf

NOTE: There were 69493 observations read from the data set PROJECT.HOUSENEW.

NOTE: The data set WORK.ONE has 69493 observations and 10 variables.

NOTE: DATA statement used (Total process time):

real time 8.50 seconds

cpu time 0.11 seconds

81 proc contents; run;

NOTE: PROCEDURE CONTENTS used (Total process time):

real time 0.21 seconds

cpu time 0.20 seconds

82 ods rtf close; **SAS OUTPUT FILE:**

01:42 Wednesday, May 20, 2020 **1**

|  |
| --- |
| ***The SAS System*** |

|  |
| --- |
| ***The CONTENTS Procedure*** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Set Name** | WORK.ONE | **Observations** | 69493 |
| **Member Type** | DATA | **Variables** | 10 |
| **Engine** | V9 | **Indexes** | 0 |
| **Created** | 05/15/2020 18:06:52 | **Observation Length** | 48 |
| **Last Modified** | 05/15/2020 18:06:52 | **Deleted Observations** | 0 |
| **Protection** |  | **Compressed** | NO |
| **Data Set Type** |  | **Sorted** | NO |
| **Label** |  |  |  |
| **Data Representation** | WINDOWS\_32 |  |  |
| **Encoding** | wlatin1 Western (Windows) |  |  |

| **Engine/Host Dependent Information** | |
| --- | --- |
| **Data Set Page Size** | 65536 |
| **Number of Data Set Pages** | 52 |
| **First Data Page** | 1 |
| **Max Obs per Page** | 1361 |
| **Obs in First Data Page** | 1315 |
| **Number of Data Set Repairs** | 0 |
| **ExtendObsCounter** | YES |
| **Filename** | C:\Users\HABIBA~1.AZI\AppData\Local\Temp\SAS Temporary Files\\_TD3088\_VDI-P2-0022\_\one.sas7bdat |
| **Release Created** | 9.0401M2 |
| **Host Created** | W32\_7PRO |

| **Alphabetic List of Variables and Attributes** | | | | |
| --- | --- | --- | --- | --- |
| **#** | **Variable** | **Type** | **Len** | **Label** |
| **2** | BLD | Char | 2 | Type of housing unit |
| **10** | FINCP | Num | 8 | Family income |
| **6** | GASAMT | Num | 8 | Monthly gas amount |
| **5** | HHAGE | Num | 3 | Age of householder |
| **8** | HHGRAD | Char | 2 | Educational level of householder |
| **4** | HHMAR | Char | 1 | Marital status of householder |
| **3** | HHSEX | Char | 1 | Sex of householder |
| **9** | HINCP | Num | 8 | Household Income |
| **1** | TENURE | Char | 1 | Owner or renter status of unit |
| **7** | UTILAMT | Num | 8 | Monthly total utility amount |

II.1 excluding all missing values and recoding few variables.

**SAS PROGRAM FILE:**

**DATA** TWO; SET one;

IF HINCP < **0** THEN HINCP = **.**;

IF FINCP < **0** THEN FINCP = **.**;

IF NMISS(HHSEX,HHMAR,HHGRAD,TENURE,BLD,UTILAMT,HHAGE,HINCP,GASAMT,FINCP) = **0**;

IF HHGRAD <=**39** THEN EDGRP = **1**;\*\*\*\* HIGH SCHOOL OR LESS;

ELSE IF **39** < HHGRAD <=**44** THEN EDGRP = **2**;\*\*\* BACHELOR'S OR LESS;

ELSE IF **44** < HHGRAD <=**47** THEN EDGRP = **3**;\*\*\* MASTERS DEGREE, PROFESSIONAL DEGREE, PHD;

IF HINCP <= **22800** THEN INGRP = **1**; \*\*\* LOWEST HOUSEHOLD INCOME GROUP;

ELSE IF **22800** < HINCP <= **50000** THEN INGRP = **2**; \*\*\* LOW HOUSEHOLD INCOME GROUP;

ELSE IF **50000** < HINCP <= **90000** THEN INGRP = **3**; \*\*\* HIGH HOUSEHOLD INCOME GROUP;

ELSE IF HINCP > **90000** THEN INGRP = **4**; \*\*\* HIGHEST HOUSEHOLD INCOME GROUP;

**PROC** **FREQ**; TABLES HHSEX HHMAR TENURE BLD EDGRP INGRP ;**RUN**;

**PROC** **MEANS**; VAR UTILAMT HHAGE GASAMT FINCP; **RUN**;

**RUN**;

**SAS LOG FILE:**

126 DATA TWO; SET one;

127

128 IF HINCP < 0 THEN HINCP = .;

129 IF FINCP < 0 THEN FINCP = .;

130 IF NMISS(HHSEX,HHMAR,HHGRAD,TENURE,BLD,UTILAMT,HHAGE,HINCP,GASAMT,FINCP) = 0;

131

132

133 IF HHGRAD <=39 THEN EDGRP = 1;\*\*\*\* HIGH SCHOOL OR LESS;

134 ELSE IF 39 < HHGRAD <=44 THEN EDGRP = 2;\*\*\* BACHELOR'S OR LESS;

135 ELSE IF 44 < HHGRAD <=47 THEN EDGRP = 3;\*\*\* MASTERS DEGREE, PROFESSIONAL DEGREE, PHD;

136

137 IF HINCP <= 22800 THEN INGRP = 1; \*\*\* LOWEST HOUSEHOLD INCOME GROUP;

138 ELSE IF 22800 < HINCP <= 50000 THEN INGRP = 2; \*\*\* LOW HOUSEHOLD INCOME GROUP;

139 ELSE IF 50000 < HINCP <= 90000 THEN INGRP = 3; \*\*\* HIGH HOUSEHOLD INCOME GROUP;

140 ELSE IF HINCP > 90000 THEN INGRP = 4; \*\*\* HIGHEST HOUSEHOLD INCOME GROUP;

141

142

143 ODS RTF;

NOTE: Writing RTF Body file: sasrtf.rtf

NOTE: Character values have been converted to numeric values at the places given by: (Line):(Column).

130:10 130:16 130:22 130:29 130:36 133:4 134:14 135:14

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=09 HHSEX=N HHMAR=N HHAGE=N GASAMT=N UTILAMT=N HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=7

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=02 HHSEX=N HHMAR=N HHAGE=N GASAMT=N UTILAMT=N HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=8

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=03 HHSEX=N HHMAR=N HHAGE=N GASAMT=1 UTILAMT=0 HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=17

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=03 HHSEX=N HHMAR=N HHAGE=N GASAMT=0 UTILAMT=0 HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=23

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=07 HHSEX=N HHMAR=N HHAGE=N GASAMT=0 UTILAMT=0 HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=27

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=06 HHSEX=N HHMAR=N HHAGE=N GASAMT=1 UTILAMT=0 HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=28

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=01 HHSEX=N HHMAR=N HHAGE=N GASAMT=1 UTILAMT=0 HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=40

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=08 HHSEX=N HHMAR=N HHAGE=N GASAMT=0 UTILAMT=0 HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=45

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=06 HHSEX=N HHMAR=N HHAGE=N GASAMT=0 UTILAMT=0 HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=53

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=03 HHSEX=N HHMAR=N HHAGE=N GASAMT=1 UTILAMT=0 HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=63

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=04 HHSEX=N HHMAR=N HHAGE=N GASAMT=N UTILAMT=N HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=73

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=02 HHSEX=N HHMAR=N HHAGE=N GASAMT=N UTILAMT=N HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=83

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=02 HHSEX=N HHMAR=N HHAGE=N GASAMT=N UTILAMT=0 HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=84

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=01 HHSEX=N HHMAR=N HHAGE=N GASAMT=N UTILAMT=N HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=87

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=02 HHSEX=N HHMAR=N HHAGE=N GASAMT=N UTILAMT=N HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=93

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=01 HHSEX=N HHMAR=N HHAGE=N GASAMT=0 UTILAMT=0 HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=102

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=06 HHSEX=N HHMAR=N HHAGE=N GASAMT=0 UTILAMT=0 HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=103

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=05 HHSEX=N HHMAR=N HHAGE=N GASAMT=2 UTILAMT=0 HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=110

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

TENURE=N BLD=04 HHSEX=N HHMAR=N HHAGE=N GASAMT=N UTILAMT=N HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=113

NOTE: Invalid numeric data, HHSEX='N' , at line 130 column 10.

NOTE: Invalid numeric data, HHMAR='N' , at line 130 column 16.

NOTE: Invalid numeric data, HHGRAD='N' , at line 130 column 22.

NOTE: Invalid numeric data, TENURE='N' , at line 130 column 29.

WARNING: Limit set by ERRORS= option reached. Further errors of this type will not be printed.

TENURE=N BLD=05 HHSEX=N HHMAR=N HHAGE=N GASAMT=0 UTILAMT=0 HHGRAD=N HINCP=. FINCP=. EDGRP=. INGRP=.

\_ERROR\_=1 \_N\_=133

NOTE: There were 69493 observations read from the data set WORK.ONE.

NOTE: The data set WORK.TWO has 60482 observations and 12 variables.

NOTE: DATA statement used (Total process time):

real time 0.15 seconds

cpu time 0.12 seconds

144 PROC FREQ; TABLES HHSEX HHMAR TENURE BLD EDGRP INGRP ;RUN;

NOTE: There were 60482 observations read from the data set WORK.TWO.

NOTE: PROCEDURE FREQ used (Total process time):

real time 0.23 seconds

cpu time 0.20 seconds

145 PROC MEANS; VAR UTILAMT HHAGE GASAMT FINCP; RUN;

NOTE: There were 60482 observations read from the data set WORK.TWO.

NOTE: PROCEDURE MEANS used (Total process time):

real time 0.09 seconds

cpu time 0.07 seconds

146 RUN;

147 ODS RTF CLOSE; **SAS OUTPUT FILE:**

01:42 Wednesday, May 20, 2020 **1**

| **Sex of householder** | | | | |
| --- | --- | --- | --- | --- |
| **HHSEX** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **1** | 30554 | 50.52 | 30554 | 50.52 |
| **2** | 29928 | 49.48 | 60482 | 100.00 |

| **Marital status of householder** | | | | |
| --- | --- | --- | --- | --- |
| **HHMAR** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **1** | 27640 | 45.70 | 27640 | 45.70 |
| **2** | 1285 | 2.12 | 28925 | 47.82 |
| **3** | 5952 | 9.84 | 34877 | 57.67 |
| **4** | 10010 | 16.55 | 44887 | 74.22 |
| **5** | 1974 | 3.26 | 46861 | 77.48 |
| **6** | 13621 | 22.52 | 60482 | 100.00 |

| **Owner or renter status of unit** | | | | |
| --- | --- | --- | --- | --- |
| **TENURE** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **1** | 35318 | 58.39 | 35318 | 58.39 |
| **2** | 24221 | 40.05 | 59539 | 98.44 |
| **3** | 943 | 1.56 | 60482 | 100.00 |

| **Type of housing unit** | | | | |
| --- | --- | --- | --- | --- |
| **BLD** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **01** | 2818 | 4.66 | 2818 | 4.66 |
| **02** | 35749 | 59.11 | 38567 | 63.77 |
| **03** | 5340 | 8.83 | 43907 | 72.60 |
| **04** | 1992 | 3.29 | 45899 | 75.89 |
| **05** | 2710 | 4.48 | 48609 | 80.37 |
| **06** | 3114 | 5.15 | 51723 | 85.52 |
| **07** | 2895 | 4.79 | 54618 | 90.30 |
| **08** | 2313 | 3.82 | 56931 | 94.13 |
| **09** | 3524 | 5.83 | 60455 | 99.96 |
| **10** | 27 | 0.04 | 60482 | 100.00 |

| **EDGRP** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **1** | 22507 | 37.21 | 22507 | 37.21 |
| **2** | 30065 | 49.71 | 52572 | 86.92 |
| **3** | 7910 | 13.08 | 60482 | 100.00 |

| **INGRP** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **1** | 15280 | 25.26 | 15280 | 25.26 |
| **2** | 15332 | 25.35 | 30612 | 50.61 |
| **3** | 13755 | 22.74 | 44367 | 73.36 |
| **4** | 16115 | 26.64 | 60482 | 100.00 |

| **Variable** | **Label** | **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| UTILAMT HHAGE GASAMT FINCP | Monthly total utility amount Age of householder Monthly gas amount Family income | 60482 60482 60482 60482 | 204.7788102 51.5303396 37.2303991 71955.43 | 120.5974597 16.6560271 42.1792744 115287.82 | 0 15.0000000 0 0 | 1010.00 85.0000000 280.0000000 5786000.00 |

II.1 draw a random sample

**SAS PROGRAM FILE:**

**DATA** THREE; SET two;

ODS RTF;

**proc** **surveyselect** method=srs n=**2000** seed=**12345** out=samplepro; **run** ;

**DATA** four; set samplepro;

**PROC** **FREQ**; TABLES HHSEX HHMAR TENURE BLD EDGRP INGRP ;**RUN**;

**PROC** **MEANS**; VAR UTILAMT HHAGE GASAMT FINCP; **RUN**;

**RUN**;

ODS RTF CLOSE;

**SAS LOG FILE:**

148 DATA THREE; SET two;

149 ODS RTF;

NOTE: Writing RTF Body file: sasrtf.rtf

NOTE: There were 60482 observations read from the data set WORK.TWO.

NOTE: The data set WORK.THREE has 60482 observations and 12 variables.

NOTE: DATA statement used (Total process time):

real time 0.12 seconds

cpu time 0.10 seconds

150 proc surveyselect method=srs n=2000 seed=12345 out=samplepro; run ;

NOTE: The data set WORK.SAMPLEPRO has 2000 observations and 12 variables.

NOTE: PROCEDURE SURVEYSELECT used (Total process time):

real time 0.12 seconds

cpu time 0.10 seconds

151

152 DATA four; set samplepro;

NOTE: There were 2000 observations read from the data set WORK.SAMPLEPRO.

NOTE: The data set WORK.FOUR has 2000 observations and 12 variables.

NOTE: DATA statement used (Total process time):

real time 0.03 seconds

cpu time 0.01 seconds

153 PROC FREQ; TABLES HHSEX HHMAR TENURE BLD EDGRP INGRP ;RUN;

NOTE: There were 2000 observations read from the data set WORK.FOUR.

NOTE: PROCEDURE FREQ used (Total process time):

real time 0.25 seconds

cpu time 0.22 seconds

154 PROC MEANS; VAR UTILAMT HHAGE GASAMT FINCP; RUN;

NOTE: There were 2000 observations read from the data set WORK.FOUR.

NOTE: PROCEDURE MEANS used (Total process time):

real time 0.10 seconds

cpu time 0.09 seconds

155 RUN;

156

157 ODS RTF CLOSE; **SAS OUTPUT FILE:**

|  |
| --- |
| ***The SAS System*** |

|  |
| --- |
| ***The SURVEYSELECT Procedure*** |

|  |  |
| --- | --- |
| **Selection Method** | Simple Random Sampling |

|  |  |
| --- | --- |
| **Input Data Set** | THREE |
| **Random Number Seed** | 12345 |
| **Sample Size** | 2000 |
| **Selection Probability** | 0.033068 |
| **Sampling Weight** | 30.241 |
| **Output Data Set** | SAMPLEPRO |

| **Sex of householder** | | | | |
| --- | --- | --- | --- | --- |
| **HHSEX** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **1** | 1000 | 50.00 | 1000 | 50.00 |
| **2** | 1000 | 50.00 | 2000 | 100.00 |

| **Marital status of householder** | | | | |
| --- | --- | --- | --- | --- |
| **HHMAR** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **1** | 951 | 47.55 | 951 | 47.55 |
| **2** | 42 | 2.10 | 993 | 49.65 |
| **3** | 174 | 8.70 | 1167 | 58.35 |
| **4** | 279 | 13.95 | 1446 | 72.30 |
| **5** | 69 | 3.45 | 1515 | 75.75 |
| **6** | 485 | 24.25 | 2000 | 100.00 |

| **Owner or renter status of unit** | | | | |
| --- | --- | --- | --- | --- |
| **TENURE** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **1** | 1194 | 59.70 | 1194 | 59.70 |
| **2** | 773 | 38.65 | 1967 | 98.35 |
| **3** | 33 | 1.65 | 2000 | 100.00 |

| **Type of housing unit** | | | | |
| --- | --- | --- | --- | --- |
| **BLD** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **01** | 76 | 3.80 | 76 | 3.80 |
| **02** | 1212 | 60.60 | 1288 | 64.40 |
| **03** | 182 | 9.10 | 1470 | 73.50 |
| **04** | 60 | 3.00 | 1530 | 76.50 |
| **05** | 80 | 4.00 | 1610 | 80.50 |
| **06** | 99 | 4.95 | 1709 | 85.45 |
| **07** | 106 | 5.30 | 1815 | 90.75 |
| **08** | 81 | 4.05 | 1896 | 94.80 |
| **09** | 101 | 5.05 | 1997 | 99.85 |
| **10** | 3 | 0.15 | 2000 | 100.00 |

| **EDGRP** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **1** | 740 | 37.00 | 740 | 37.00 |
| **2** | 985 | 49.25 | 1725 | 86.25 |
| **3** | 275 | 13.75 | 2000 | 100.00 |

| **INGRP** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **1** | 469 | 23.45 | 469 | 23.45 |
| **2** | 500 | 25.00 | 969 | 48.45 |
| **3** | 491 | 24.55 | 1460 | 73.00 |
| **4** | 540 | 27.00 | 2000 | 100.00 |

| **Variable** | **Label** | **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| UTILAMT HHAGE GASAMT FINCP | Monthly total utility amount Age of householder Monthly gas amount Family income | 2000 2000 2000 2000 | 205.5410000 50.7610000 37.2070000 74189.85 | 116.7102650 16.6814572 40.9338606 121388.64 | 0 17.0000000 0 0 | 790.0000000 85.0000000 250.0000000 2934700.00 |

III.1. Build confidence intervals

**SAS PROGRAM FILE:**

\*\*\* Building CI;

\*\*\* I/R variable;

**DATA** five; set samplepro;

ods rtf;

**proc** **means** maxdec=**2** n mean stddev clm alpha=**0.05** ;

var UTILAMT ;

**run** ;

\*\*\* Proportion;

**PROC** **SURVEYFREQ**; TABLES HHSEX/CL;

**RUN**;

ods rtf close;

**SAS LOG FILE:**

36 \*\*\* Building CI;

37

38 \*\*\* I/R variable;

39 DATA five; set samplepro;

40 ods rtf;

NOTE: Writing RTF Body file: sasrtf.rtf

NOTE: There were 2000 observations read from the data set WORK.SAMPLEPRO.

NOTE: The data set WORK.FIVE has 2000 observations and 12 variables.

NOTE: DATA statement used (Total process time):

real time 0.31 seconds

cpu time 0.07 seconds

41 proc means maxdec=2 n mean stddev clm alpha=0.05 ;

42 var UTILAMT ;

43 run ;

NOTE: There were 2000 observations read from the data set WORK.FIVE.

NOTE: PROCEDURE MEANS used (Total process time):

real time 0.12 seconds

cpu time 0.07 seconds

44 \*\*\* Proportion;

45 PROC SURVEYFREQ; TABLES HHSEX/CL;

46 RUN;

NOTE: PROCEDURE SURVEYFREQ used (Total process time):

real time 0.12 seconds

cpu time 0.04 seconds

47

48 ods rtf close; **SAS OUTPUT FILE:**

| **Analysis Variable : UTILAMT Monthly total utility amount** | | | | |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **Std Dev** | **Lower 95% CL for Mean** | **Upper 95% CL for Mean** |
| 2000 | 205.54 | 116.71 | 200.42 | 210.66 |

| **Data Summary** | |
| --- | --- |
| **Number of Observations** | 2000 |

| **Sex of householder** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **HHSEX** | **Frequency** | **Percent** | **Std Err of Percent** | **95% Confidence Limits for Percent** | |
| **1** | 1000 | 50.0000 | 1.1183 | 47.8068 | 52.1932 |
| **2** | 1000 | 50.0000 | 1.1183 | 47.8068 | 52.1932 |
|  |  |  |  |  |  |
| **Total** | 2000 | 100.000 |  |  |  |

III.2. one- sample t-tests

**SAS PROGRAM FILE:**

**Data** six; set samplepro;

ODS RTF;

Title 'One-sample t-test using PROC TTEST';

**PROC** **TTEST** H0=**51.5303396**; var HHAGE;

**RUN** ;

ODS RTF CLOSE;

**SAS LOG FILE:**

54 Data six; set samplepro;

55 ODS RTF;

56 Title 'One-sample t-test using PROC TTEST';

NOTE: There were 2000 observations read from the data set

WORK.SAMPLEPRO.

NOTE: The data set WORK.SIX has 2000 observations and 12

variables.

NOTE: DATA statement used (Total process time):

real time 0.03 seconds

cpu time 0.03 seconds

57 PROC TTEST H0=51.5303396; var HHAGE;

58 RUN ;

NOTE: PROCEDURE TTEST used (Total process time):

real time 2.81 seconds

cpu time 1.03 seconds

59 ODS RTF CLOSE;

**SAS OUTPUT FILE:**

|  |
| --- |
| ***One-sample t-test using PROC TTEST*** |

|  |
| --- |
| ***The TTEST Procedure*** |

|  |
| --- |
| ***Variable: HHAGE (Age of householder)*** |

| **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| 2000 | 50.7610 | 16.6815 | 0.3730 | 17.0000 | 85.0000 |

| **Mean** | **95% CL Mean** | | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- | --- |
| 50.7610 | 50.0295 | 51.4925 | 16.6815 | | 16.1801 | 17.2152 |

| **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- |
| 1999 | -2.06 | 0.0393 |





III.2. two-sample t-tests

**SAS PROGRAM FILE:**

**data** seven; set samplepro;

ods rtf;

**PROC** **FREQ**; TABLES HHSEX; **RUN**;

**PROC** **MEANS**; VAR FINCP; **RUN**;

**PROC** **TTEST** ;

CLASS HHSEX; VAR FINCP; **RUN**;

Title 'Independent Group t-Test';

**RUN**;

ods rtf close;

**SAS LOG FILE:**

60 data seven; set samplepro;

61 ods rtf;

NOTE: Writing RTF Body file: sasrtf.rtf

NOTE: There were 2000 observations read from the data set WORK.SAMPLEPRO.

NOTE: The data set WORK.SEVEN has 2000 observations and 12 variables.

NOTE: DATA statement used (Total process time):

real time 0.14 seconds

cpu time 0.10 seconds

62 PROC FREQ; TABLES HHSEX; RUN;

NOTE: There were 2000 observations read from the data set WORK.SEVEN.

NOTE: PROCEDURE FREQ used (Total process time):

real time 0.07 seconds

cpu time 0.06 seconds

63 PROC MEANS; VAR FINCP; RUN;

NOTE: There were 2000 observations read from the data set WORK.SEVEN.

NOTE: PROCEDURE MEANS used (Total process time):

real time 0.07 seconds

cpu time 0.06 seconds

64

65 PROC TTEST ;

66 CLASS HHSEX; VAR FINCP; RUN;

NOTE: PROCEDURE TTEST used (Total process time):

real time 3.62 seconds

cpu time 1.50 seconds

67 Title 'Independent Group t-Test';

68 RUN;

69 ods rtf close; **SAS OUTPUT FILE:**

| **Sex of householder** | | | | |
| --- | --- | --- | --- | --- |
| **HHSEX** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **1** | 1000 | 50.00 | 1000 | 50.00 |
| **2** | 1000 | 50.00 | 2000 | 100.00 |

| **Analysis Variable : FINCP Family income** | | | | |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| 2000 | 74189.85 | 121388.64 | 0 | 2934700.00 |

| **HHSEX** | **N** | **Mean** | **Std Dev** | **Std Err** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | 1000 | 85968.2 | 140057 | 4429.0 | 0 | 2934700 |
| **2** | 1000 | 62411.5 | 97936.2 | 3097.0 | 0 | 1709000 |
| **Diff (1-2)** |  | 23556.7 | 120846 | 5404.4 |  |  |

| **HHSEX** | **Method** | **Mean** | **95% CL Mean** | | **Std Dev** | **95% CL Std Dev** | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **1** |  | 85968.2 | 77277.0 | 94659.4 | 140057 | 134176 | 146481 |
| **2** |  | 62411.5 | 56334.1 | 68488.9 | 97936.2 | 93824.1 | 102428 |
| **Diff (1-2)** | **Pooled** | 23556.7 | 12957.8 | 34155.5 | 120846 | 117213 | 124713 |
| **Diff (1-2)** | **Satterthwaite** | 23556.7 | 12957.1 | 34156.2 |  |  |  |

| **Method** | **Variances** | **DF** | **t Value** | **Pr > |t|** |
| --- | --- | --- | --- | --- |
| **Pooled** | Equal | 1998 | 4.36 | <.0001 |
| **Satterthwaite** | Unequal | 1787.4 | 4.36 | <.0001 |

| **Equality of Variances** | | | | |
| --- | --- | --- | --- | --- |
| **Method** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **Folded F** | 999 | 999 | 2.05 | <.0001 |





III.3. Use Chi-square test.

**SAS PROGRAM FILE:**

**data** eight; set samplepro;

\*\*\* tenure will be recoded in 3 groups;

if tenure = **1** then hhgroup=**1**;

if tenure = **2** then hhgroup=**2**;

if tenure = **3** then hhgroup=**3**;

ods rtf;

**proc** **freq**; tables hhgroup\*hhsex/chisq; **run**;

ods rtf close;

**SAS LOG FILE:**

98 data eight; set samplepro;

99 \*\*\* tenure will be recoded in 3 groups;

100 if tenure = 1 then hhgroup=1;

101 if tenure = 2 then hhgroup=2;

102 if tenure = 3 then hhgroup=3;

103

104 ods rtf;

NOTE: Writing RTF Body file: sasrtf.rtf

NOTE: Character values have been converted to numeric

values at the places given by: (Line):(Column).

100:4 101:4 102:4

NOTE: There were 2000 observations read from the data set WORK.SAMPLEPRO.

NOTE: The data set WORK.EIGHT has 2000 observations and 13 variables.

NOTE: DATA statement used (Total process time):

real time 0.12 seconds

cpu time 0.09 seconds

105 proc freq; tables hhgroup\*hhsex/chisq; run;

NOTE: There were 2000 observations read from the data set WORK.EIGHT.

NOTE: PROCEDURE FREQ used (Total process time):

real time 0.20 seconds

cpu time 0.18 seconds

106

107 ods rtf close;

**SAS OUTPUT FILE:**

| **Table of hhgroup by HHSEX** | | | |
| --- | --- | --- | --- |
| **hhgroup** | **HHSEX(Sex of householder)** | | |
| **Frequency Percent Row Pct Col Pct** | **1** | **2** | **Total** |
| **1** | 634 31.70 53.10 63.40 | 560 28.00 46.90 56.00 | 1194 59.70 |
| **2** | 352 17.60 45.54 35.20 | 421 21.05 54.46 42.10 | 773 38.65 |
| **3** | 14 0.70 42.42 1.40 | 19 0.95 57.58 1.90 | 33 1.65 |
| **Total** | 1000 50.00 | 1000 50.00 | 2000 100.00 |

|  |  |  |
| --- | --- | --- |
| ***Statistics for Table of hhgroup by HHSEX*** | | |
|  | | |
| **Statistic** | | **DF** | **Value** | | **Prob** |
| **Chi-Square** | | 2 | 11.5030 | | 0.0032 |
| **Likelihood Ratio Chi-Square** | | 2 | 11.5170 | | 0.0032 |
| **Mantel-Haenszel Chi-Square** | | 1 | 11.2793 | | 0.0008 |
| **Phi Coefficient** | |  | 0.0758 | |  |
| **Contingency Coefficient** | |  | 0.0756 | |  |
| **Cramer's V** | |  | 0.0758 | |  |

***Sample Size = 2000***

III.4. Crosstabulation analysis

**SAS PROGRAM FILE:**

**data** nine; set samplepro;

\*\*\* Income group is already recoded into 4 categories from HINCP i.e., from lowest to highest in DATA TWO step;

\*\*\* EDGRP i.e., HHGRAD is also recode in 3 categories in DATA two step;

\*\*\* Showing the recoding step from data two in comments below;

\*\*\*IF HINCP <= 22800 THEN INGRP = 1; \*\*\* LOWEST HOUSEHOLD INCOME GROUP;

\*\*\*ELSE IF 22800 < HINCP <= 50000 THEN INGRP = 2; \*\*\* LOW HOUSEHOLD INCOME GROUP;

\*\*\*ELSE IF 50000 < HINCP <= 90000 THEN INGRP = 3; \*\*\* HIGH HOUSEHOLD INCOME GROUP;

\*\*\*ELSE IF HINCP > 90000 THEN INGRP = 4; \*\*\* HIGHEST HOUSEHOLD INCOME GROUP;

\*\*\*IF HHGRAD <=39 THEN EDGRP = 1;\*\*\*\* HIGH SCHOOL OR LESS;

\*\*\*ELSE IF 39 < HHGRAD <=44 THEN EDGRP = 2;\*\*\* BACHELOR'S OR LESS;

\*\*\*ELSE IF 44 < HHGRAD <=47 THEN EDGRP = 3;\*\*\* MASTERS DEGREE, PROFESSIONAL DEGREE, PHD;

ODS RTF;

**PROC** **FREQ**; TABLES INGRP\*EDGRP;

**RUN**;

ODS RTF CLOSE;

**SAS LOG FILE:**

166 data nine; set samplepro;

167

168 \*\*\* Income group is already recoded into 4 categories from HINCP i.e.,

168! from lowest to highest in DATA TWO step;

169 \*\*\* EDGRP i.e., HHGRAD is also recode in 3 categories in DATA two step;

170

171 \*\*\* Showing the recoding step from data two in comments below;

172

173 \*\*\*IF HINCP <= 22800 THEN INGRP = 1; \*\*\* LOWEST HOUSEHOLD INCOME GROUP;

174 \*\*\*ELSE IF 22800 < HINCP <= 50000 THEN INGRP = 2; \*\*\* LOW HOUSEHOLD INCOME

174! GROUP;

175 \*\*\*ELSE IF 50000 < HINCP <= 90000 THEN INGRP = 3; \*\*\* HIGH HOUSEHOLD

175! INCOME GROUP;

176 \*\*\*ELSE IF HINCP > 90000 THEN INGRP = 4; \*\*\* HIGHEST HOUSEHOLD INCOME

176! GROUP;

177

178 \*\*\*IF HHGRAD <=39 THEN EDGRP = 1;\*\*\*\* HIGH SCHOOL OR LESS;

179 \*\*\*ELSE IF 39 < HHGRAD <=44 THEN EDGRP = 2;\*\*\* BACHELOR'S OR LESS;

180 \*\*\*ELSE IF 44 < HHGRAD <=47 THEN EDGRP = 3;\*\*\* MASTERS DEGREE,

180! PROFESSIONAL DEGREE, PHD;

181

182 ODS RTF;

NOTE: Writing RTF Body file: sasrtf.rtf

NOTE: There were 2000 observations read from the data set WORK.SAMPLEPRO.

NOTE: The data set WORK.NINE has 2000 observations and 12 variables.

NOTE: DATA statement used (Total process time):

real time 0.10 seconds

cpu time 0.07 seconds

183 PROC FREQ; TABLES INGRP\*EDGRP;

184 RUN;

NOTE: There were 2000 observations read from the data set WORK.NINE.

NOTE: PROCEDURE FREQ used (Total process time):

real time 0.14 seconds

cpu time 0.12 seconds

185 ODS RTF CLOSE; **SAS OUTPUT FILE:**

| **Table of INGRP by EDGRP** | | | | |
| --- | --- | --- | --- | --- |
| **INGRP** | **EDGRP** | | | |
| **Frequency Percent Row Pct Col Pct** | **1** | **2** | **3** | **Total** |
| **1** | 277 13.85 59.06 37.43 | 175 8.75 37.31 17.77 | 17 0.85 3.62 6.18 | 469 23.45 |
| **2** | 229 11.45 45.80 30.95 | 238 11.90 47.60 24.16 | 33 1.65 6.60 12.00 | 500 25.00 |
| **3** | 154 7.70 31.36 20.81 | 271 13.55 55.19 27.51 | 66 3.30 13.44 24.00 | 491 24.55 |
| **4** | 80 4.00 14.81 10.81 | 301 15.05 55.74 30.56 | 159 7.95 29.44 57.82 | 540 27.00 |
| **Total** | 740 37.00 | 985 49.25 | 275 13.75 | 2000 100.00 |

III.5. One-way Analysis of Variance

**SAS PROGRAM FILE:**

**data** nine; set samplepro;

IF HHMAR LE **2** THEN MARIGROUP=**1**; \*\*\*MARRIED;

IF HHMAR GT **2** AND MAR LE **6** THEN MARIGROUP=**2**; \*\*\*MARRIED and spouse absent, WIDOWED,DIVORCED, SEPARATED;

IF HHMAR = **6** THEN MARIGROUP=**3**; \*\*\*NEVER MARRIED;

ODS RTF;

**PROC** **ANOVA**;

CLASS MARIGROUP;

MODEL UTILAMT=MARIGROUP;

MEANS MARIGROUP/TUKEY ;

**RUN**;

ODS RTF CLOSE;

**SAS LOG FILE:**

NOTE: PROCEDURE ANOVA used (Total process time):

real time 36.99 seconds

cpu time 1.59 seconds

210 data nine; set samplepro;

211

212 IF HHMAR LE 2 THEN MARIGROUP=1; \*\*\*MARRIED;

213 IF HHMAR GT 2 AND MAR LE 6 THEN MARIGROUP=2; \*\*\*MARRIED and

213! spouse absent, WIDOWED,DIVORCED, SEPARATED;

214 IF HHMAR = 6 THEN MARIGROUP=3; \*\*\*NEVER MARRIED;

215

216 ODS RTF;

NOTE: Character values have been converted to numeric

values at the places given by: (Line):(Column).

212:4 213:4 214:4

NOTE: Variable MAR is uninitialized.

NOTE: There were 2000 observations read from the data set

WORK.SAMPLEPRO.

NOTE: The data set WORK.NINE has 2000 observations and 14

variables.

NOTE: DATA statement used (Total process time):

real time 0.04 seconds

cpu time 0.03 seconds

217 PROC ANOVA;

218 CLASS MARIGROUP;

219 MODEL UTILAMT=MARIGROUP;

220 MEANS MARIGROUP/TUKEY ;

221 RUN;

222 ODS RTF CLOSE; **SAS OUTPUT FILE:**

| **Class Level Information** | | |
| --- | --- | --- |
| **Class** | **Levels** | **Values** |
| **MARIGROUP** | 3 | 1 2 3 |

|  |  |
| --- | --- |
| **Number of Observations Read** | 2000 |
| **Number of Observations Used** | 2000 |

|  |
| --- |
| ***Independent Group t-Test*** |

|  |
| --- |
| ***The ANOVA Procedure*** |

|  |
| --- |
| ***Dependent Variable: UTILAMT Monthly total utility amount*** |

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 2 | 4366701.53 | 2183350.76 | 190.71 | <.0001 |
| **Error** | 1997 | 22862249.11 | 11448.30 |  |  |
| **Corrected Total** | 1999 | 27228950.64 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **UTILAMT Mean** |
| --- | --- | --- | --- |
| 0.160370 | 52.05614 | 106.9967 | 205.5410 |

| **Source** | **DF** | **Anova SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **MARIGROUP** | 2 | 4366701.526 | 2183350.763 | 190.71 | <.0001 |





|  |
| --- |
| ***Tukey's Studentized Range (HSD) Test for UTILAMT*** |

|  |  |
| --- | --- |
| **Alpha** | 0.05 |
| **Error Degrees of Freedom** | 1997 |
| **Error Mean Square** | 11448.3 |
| **Critical Value of Studentized Range** | 3.31697 |

| **Comparisons significant at the 0.05 level are indicated by \*\*\*.** | | | | |
| --- | --- | --- | --- | --- |
| **MARIGROUP Comparison** | **Difference Between Means** | **Simultaneous 95% Confidence Limits** | |  |
| **1 - 2** | 67.144 | 53.577 | 80.712 | \*\*\* |
| **1 - 3** | 110.812 | 96.909 | 124.714 | \*\*\* |
| **2 - 1** | -67.144 | -80.712 | -53.577 | \*\*\* |
| **2 - 3** | 43.667 | 27.840 | 59.495 | \*\*\* |
| **3 - 1** | -110.812 | -124.714 | -96.909 | \*\*\* |
| **3 - 2** | -43.667 | -59.495 | -27.840 | \*\*\* |